DOCUMENT RESUME

ED 178 588 TM 009 778

AUTHOR Burrill, Lois F.
TITLE How a Standardized Achievement Test is Built. Test Service Notebook 125.
INSTITUTION Psychological Corp., New York, N.Y.
PUB DATE [70]
NOTE 7p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; Curriculum; Educational Objectives; Elementary Education; Field Studies; *Item Analysis; *National Norms; *Norm Referenced Tests; Sampling; *Standardized Tests; *Test Construction; Test Reliability; Test Validity
IDENTIFIERS Harcourt Brace Jovanovich; *Testing Industry; Test Manuals

ABSTRACT

Procedures which are used by Harcourt, Brace Jovanovich, Inc., in developing achievement tests are described; these steps are suggested to be typical of major test publishers. A curricular analysis of fifth grade mathematics textbooks is presented, to illustrate the process which determines which skills and objectives are tested. The characteristics of acceptable test questions are briefly described, as are item banks and considerations regarding the physical format of the test. The national item analysis program is described as the trial administration of experimental test forms, which examines item difficulty, item discrimination, and grade progression in difficulty level. Procedures for norming are described, including selection of the standardization sample; and three types of norms are defined: grade equivalents, percentile ranks, and stanines. The test manual is said to provide information to the test user concerning test reliability, test validity, and standard error of measurement. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED178588

Test Service Notebook 125

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Marilyn Lucas
Psychological Corp.
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

How a Standardized Achievement Test Is Built

LOIS E. BURRILL, *Test Department*

* * * * *

There is no one, "hard-and-fast" set of rules for building a standardized achievement test series. Each step described in this notebook is typical of the way tests are built by many of the major test publishers, but there are many variations possible in technique. In general, the text which follows describes the development of standardized achievement tests at Harcourt Brace Jovanovich, Inc.

* * * * *

Let us assume a test publisher has decided to build a new series of achievement tests for use in elementary schools. Achievement tests are designed to measure the extent to which pupils have "achieved" or mastered the skills and knowledge which are the goals of the school's instructional program. The school should then be able to use the test results in planning an instructional program to meet the needs of each pupil. Since achievement tests are so closely related to curriculum, the publisher will build his tests to reflect what is typically being taught in classrooms around the country.

Development of Experimental Forms

An essential prelude, then, to determining what the content of the test will be is the identification of what is being taught country-wide: in other words, an analysis of the curriculum. Curricular analysis for a major achievement series is a long and arduous task. It requires the detailed summarizing of published textbook series, syllabuses, outlines of objectives, and other current curricular material from across the country. (See Figure 1.) In addition to analyses of current texts and course outlines, curriculum experts in each of the skills and subject matter areas to be tested are consulted for suggestions as to what trends the curriculum might be taking in the future. Since curricular analysis and prepublication research typically extend over a period of several years, it is essential for the test publisher to make an evaluation of future curriculum trends. (The transition from "traditional" to "modern" mathematics during the sixties is an example of the type of curricular change the test developer must anticipate in his determination of what content should be included.)

Topic/Text	A	B	C	D	E
Fractions	23.4	29.0	26.4	23.4	19.4
Geometry	11.5	19.0	10.3	11.5	5.9
Rates & Ratios	13.2	2.0	2.6	12.3	2.5
Decimals	5.8	1.3	3.9	5.8	8.4
Graphs & Scale Drawings	2.7	4.0	4.8	2.7	4.4
Place Value	1.5	—	2.3	1.5	3.4
Measures	1.2	—	2.6	1.2	1.6
Other Number Bases	—	—	1.3	—	2.8

Figure 1. Part of a Topical Analysis of 5th Grade Mathematics Texts. A, B, C, etc. refer to various well-known textbook series. The number entries are the percent of total pages devoted to the topic. Taken from Curricular Analysis for *Metropolitan Achievement Tests*, 1970 Edition.

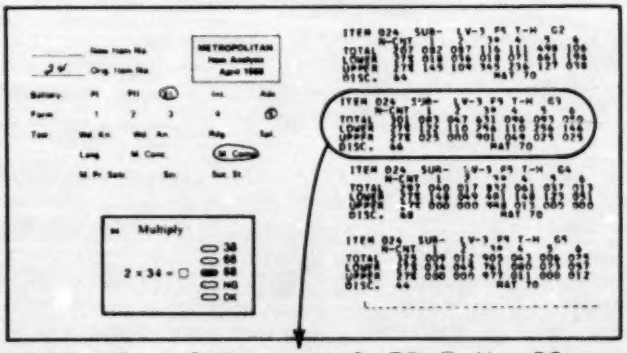
The end result of this process is a set of content outlines specifying behavioral objectives in each subject and skill area. The relative emphasis given to each objective is also analyzed so as to provide an indication of the percent of the test questions which should be developed to measure each one. The test developer will use this information as a blueprint in planning the building of the test. Often, these blueprints will be sent to curriculum experts for their review and comment.

TM009 778

The next step in the process is the actual writing of the test questions or items. Questions may be written by test authors, by teachers whom they employ to write questions, and/or by members of the publisher's staff. Item writing is not a simple process. The question, or stem, must state clearly what is being asked; the correct answer must be clear and unambiguous, and the incorrect choices (distractors) must be both attractive (to the examinee who *does not* know the answer) and clearly incorrect (to the examinee who *does* know the answer). Before ever being tried out, each test question may go through several cycles of editing, rewriting, and review by different experts. Curriculum specialists will study the questions to check that they measure objectives being taught in the classroom. Test construction experts will review them to be sure questions conform to rules of good item-writing. Other experts will review questions to determine that the context of the question does not bias its validity for any particular subgroup of the population and to make sure they will not psychologically or emotionally "turn off" a youngster. Any drawings to be used in test questions must be planned, executed, and (particularly at the lower grade levels) carefully edited and reviewed right along with the test questions to be certain that all the objectives of the question are being met.

All questions in a standardized test will be "tried out" before being included in the final version of the test. Test publishers know from past experience that, no matter how thorough the reviewing process, a number of proposed questions will have to be discarded at that time for any of a number of reasons which will be discussed below. At HBJ, editors also want to have a surplus of usable questions so that they can choose the best ones from among them. The total number of questions written, therefore, may be as many as three to four times the number of questions needed for the final versions of the test.

Many other details are discussed at this time. For example, decisions are made concerning the directions to the teacher and to the pupil. Sample questions will be written to acquaint the pupil with the task he will be asked to perform. Methods of indicating responses, type of answer documents, size of type, format of booklets, and such topics will be discussed.



ITEM 024 SUB- LV-3 F5 T-H G3

	N-CNT	1	2	3*	4	5	6
TOTAL	301	083	047	631	096	093	050
LOWER	27%	122	110	256	110	256	146
UPPER	27%	025	000	901	049	025	025
DISC.	66			MAT 70			

Figure 2. Item Card showing Data from the Item Analysis Research Program. 63.1 percent of the pupils in Grade 3 answered this question correctly. Of the top 27 percent group, 90.1 percent got the correct answer; 25.6 percent of the bottom group also answered correctly. 9.3 percent of the total group marked "Don't Know," and 5 percent omitted the item. (Note the total percent correct for each grade 2-5, the grade progression of these percents correct, and the discrimination values for each grade. This is a pretty good item.)

After the many questions have been written and edited, they are assembled into experimental test forms for "tryout" and subsequent analysis. Typically, each of these experimental forms looks like a final test form and meets the specifications set up for the test in the blueprint. Often, however, each form has more items per objective than the final form is planned to have, and some or all forms may include sets of questions experimenting with a new "item type" or a new set of directions. In page size, layout, typography, directions for administration, etc., these forms will, however, be virtually identical with those planned for the final edition, since all these aspects of the test must also be "tried out" along with the items before final decisions are made.

Tryout: National Item Analysis Program

Tryout administration of these experimental forms is called an "item analysis" program. A publisher such as HBJ plans a research program in order to get a variety of information about the questions that have been developed for the test:

1. the difficulty level of each question – that is, what percent of the pupils in the total tryout group answer the question correctly.
2. the discrimination of each question – that is, how well the answering of this question correctly distinguishes between students who score high (within the top 27% of the sample) on the total test and those who score low (within the bottom 27% of the sample).
3. the grade progression in difficulty – that is, if a question is used in tests for several successive grades, do a progressively greater percent of pupils answer the question correctly at successively higher grades?

In addition to the statistical characteristics of individual questions, other types of information are obtained during the course of a good item analysis program. For example, teachers administering the experimental forms are requested to complete questionnaires asking their judgment concerning the appropriateness of content for their own classroom, the clarity of directions and questions, and other pertinent matters regarding the test.

The schools participating in an item analysis program are carefully chosen to be representative of the population for whom the test is designed. The communities are typically selected on the basis of such factors as size, geographical location, and socioeconomic level (as indicated by adult level of income and education, previous test results for pupils, etc.). A test of scholastic aptitude or mental ability is often administered as part of this achievement test tryout program in order to check on the ability levels represented by the participating schools, and to provide a basis for matching the groups taking the various experimental forms.

Wherever possible, classroom teachers themselves administer the tests so that the administration situation will correspond as closely as possible with that in which the final published tests will be used. However, in a tryout administration, the tests are typically administered without time limits so as to give each pupil ample opportunity to attempt all items on the experimental form. Teachers record the time required for their classes so that final time limits can be determined in such a way as to make certain that every pupil will be able to finish the test in the allotted time. Experimental forms are also often administered in the grade just above and the one just below those for which the questions are ultimately intended, in order to secure further data on the questions.

Development of Final Forms

The selection of test items for the final forms of the tests is begun only after the results of the item analysis program have been carefully studied and interpreted by the team of authors and editors. (See Figure 2.) The final selection of these questions depends jointly on content specifications and statistical requirements. A number of general statistical guidelines for selection are usually set up for use in all tests in the series. The following set of guidelines was used for a recently-published achievement series:

1. average difficulty level – the median difficulty level for the questions in the final test should be .55 for the target grade level (an average of 55% of pupils answering an item correctly).
2. range of difficulty values – the various questions in the final form should range in difficulty from .90 to .20 for the pupils in the grade for which the test is intended.
3. item discrimination – except for very easy or very difficult questions, the items in the final form must distinguish clearly between high- and low-scoring pupils.
4. grade progression – the difficulty level of each question for each grade in which it was tested should indicate an increasingly larger percent correct, grade by grade.

The questions which have survived the tryout administration phase are again reviewed by outside experts who are familiar with the needs and styles of pupils from a variety of minority backgrounds in an attempt to make sure that no questions are inappropriate or offensive to such children. Then the selected items are divided into the various final forms in such a way as to make each form match both the content and statistical specifications previously agreed upon. Both careful development of content specifications and tryout of all proposed questions are vital to the building of a test that will be worth standardizing.

The procedure described above could be applied, with some modifications, to many test-building tasks, whether or not norms were to be delevoped for the instrument. However, in order to complete the development of a *standardized* test, several further steps must be taken. Part of a typical definition of a standardized test would read something like ". . . and interpreted in reference to certain normative information." The development of those norm data is the next step in the building of a standardized test.

Norming: The National Standardization Program

Norms provide the means for comparing the performance of one pupil or group of pupils with that of some particular reference group. There are many types of reference groups which might be used. Norms may be developed by age or sex, for particular geographical areas, for private schools, for suburban areas, or for other distinct groups. To develop a norm, however, the norm group must be carefully defined.

Let us suppose that our hypothetical test publisher has determined that the group for which his test is being developed is the total elementary school population. Therefore, the most important norm data he must collect is for that reference group. He has already collected preliminary data from one sample of the national school population in his item analysis. Now, by a process called standardization, he will develop national normative information for the final forms of the test. This research program will yield data on the performance of a sample of pupils selected to represent the total population of pupils. When the sample has been well-chosen, one can assert with reasonable confidence that the norms developed from these data truly reflect the performance of the national school population in all its variety. The quality of the job the test developer does in selecting this sample population and in describing it to the user will, therefore, have tremendous long-range effects on the confidence with which test users can apply the norms.

Many variables are used to select these "reference group" samples, and later to describe them. They include characteristics of pupils themselves (chronological age, sex, mental ability scores, etc.), characteristics of school systems (public-private, teachers' salaries, time spent teaching various subjects and skills, etc.), and characteristics of the communities which the schools serve (geographical area, size of city, level of adult schooling and income, etc.). Clearly, the characteristics that are used as criteria in sampling should be those which have been shown to be related to test performance, and on which information is reasonably accessible, both to test developer and to test user. For example, Harcourt Brace Jovanovich, Inc., in selecting school systems for its standardization samples, uses the following community variables:

1. size of community
2. geographic region
3. socioeconomic index (the community's median family income and median years of schooling for persons over 24 years, combined and classified in some manner)

Using national census data for each of these variables, the test developer will draw up a model to indicate how many sample cases from communities of various types should be included in the total so as to reflect the makeup of the total national school population. School systems, chosen randomly from a list of all schools, are classified with respect to each of these variables and assigned to a cell. Certain school systems then are invited to participate in the standardization. These systems are selected in such a way as to produce a representative sample of all school systems in the nation. Other similarly classified school systems are selected as alternates to fill needed slots whenever a chosen system is unable or unwilling to participate. Schools agreeing to inclusion in the sample will be asked to provide many other pieces of information which will be useful in describing the sample population. Included in such questionnaires may be items regarding teachers' salaries and experience, promotion and grouping policies, length of school year, entrance age, availability of kindergarten experience, textbooks in use, tests used in the system, etc.

The test developer may also set up criteria for his norm group in terms of individual pupil characteristics – usually chronological age and mental ability, as measured by some standardized instrument. Although he cannot get more than a rough indication of these attributes before his actual testing begins, he will adjust his sample afterwards, as is illustrated in the following example. Frequently, a criterion for the norm group is that the mental ability of the pupils match the mental ability of the national population, both in average ability and the spread above and below average. This is usually checked by administering, along with the achievement test being standardized, an already-standardized mental ability test. The data thus collected will be analyzed to determine the median mental ability, as well as the variability of ability scores for the sample. Some adjustments may have to be made, either by eliminating cases or by weighting the results of various school systems in order to correct any errors in the original selection of the sample and to make sure that the total national school population is in fact accurately represented in this standardization sample. If the original design was well planned, little adjustment will be necessary.

The development of the norms themselves is one of the most important aspects of standardized test construction. From the scores of the pupils in the norm sample, distributions of raw scores on each level or "battery" of the tests are made. To give meaning to these scores and to provide national bench-marks, various kinds of national norms may be developed. For elementary achievement tests, the types most often developed include grade equivalents (GEs), percentile ranks (PRs), and stanines (Ss). The Harcourt Brace Jovanovich Test Department usually provides all three types of national norms for its new tests. The following sections briefly describe each type of norm and the way in which it is developed.

The development of grade-equivalent norms is quite complicated and several steps are required. First, in order to be able to translate raw scores from several different batteries or levels of the test into a single scale of grade equivalents, the raw scores on each level of the test must be related to the raw scores on the other levels. To accomplish this, pairs of tests are given to special, experimental groups and the equivalent raw scores from one test to the next are determined. In this way, raw scores on all levels can be translated into a single set of numbers. These numbers are then marked along the vertical axis of a graph, with the school years marked along the horizontal axis in months (1.0-9.9). Next, the median (middle) scores from each grade level of the standardization sample are plotted on the graph and the straight lines connecting each pair of points are smoothed out into a curved line. (See Figure 3.)

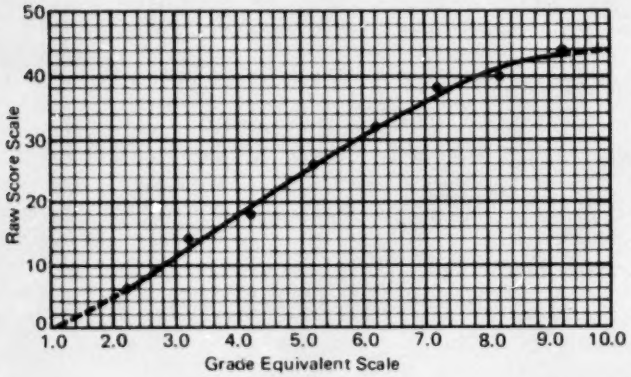


Figure 3. Schematic Illustration of Grade Equivalent Norm Line for One Subtest, Grades 2.2 - 9.2

From this norm line, then, the grade equivalent scores for each possible test score can be read and tabled. It must be remembered, however, that the only grade equivalents which are actually derived from the standardization testing are those for the one time of year at which the test was standardized: i.e., 3.2, 4.2, 5.2, etc. The other grade equivalents are estimates, taken from the smoothed curve between the obtained grade equivalents. This "interpolation" of the in-between grade equivalents involves the assumption that growth is evenly spread throughout the school year.

This explanation of how the grade equivalent norms are developed may also help explain the meaning of the grade equivalent. It is a way of expressing how the raw score of a pupil taking one test compares with the scores obtained by the *average pupils at each grade level* in the standardization administration of all levels of the test. If a pupil achieves a raw score equal to a grade equivalent of 5.8, that means his *score* was similar to that of the average 5th grader in the 8th month of the 5th grade. From the explanation given above, however, it also can be seen that a 3rd grader, taking the test designed for the 3rd grade, may have answered a quite different set of questions from the set answered by the 5th grader taking a 5th grade set of questions, although both achieve the same grade equivalent score.

Percentile rank norms provide quite a different interpretation of test scores. Instead of comparing a pupil with average pupils over a whole range of grade levels, his score is compared only with those of others at his own approximate grade placement. A percentile rank for a given raw score indicates the percent of pupils at a particular grade placement in the norm group who received scores equal to or lower than the given raw score. Developing percentile rank norms for the time of year the test was standardized is a relatively simple matter. The distribution of raw scores for the grade is plotted on a normal percentile chart, a smooth curve is fitted to the points, and the PR for each raw score is read off. (See Figure 4.) Our

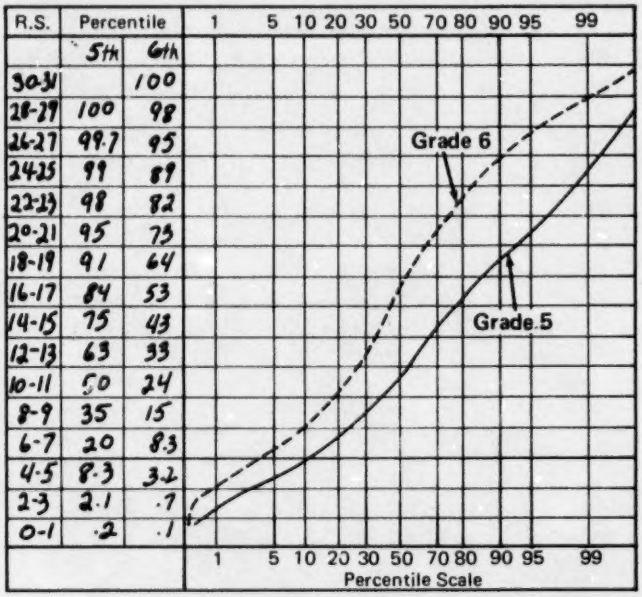


Figure 4. Schematic Illustration of Percentile Rank Norm Lines for One Subtest, Grades 5.1 and 6.1

test developer could let it go at that and provide percentile rank norms only for the time of year the test was standardized. However, he will probably want to provide at least two, perhaps three, sets of norms for various times of the school year (beginning, middle, and end-of-year, for example). Unless he standardizes his test at all of these times and with very carefully matched groups, he will have to "interpolate" norms for times of the year other than those at which he actually standardized. This is not complicated, but it involves again the assumption that growth proceeds at a constant rate throughout the school year. (See Figure 5.)

Stanine norms make use of a nine-point scale of standard scores, with a mean of 5 and a standard deviation of 2. For a normal distribution of scores such as those obtained in a national standardization, the correspondence between percentile ranks and stanines is fixed. For such distributions, the development of stanines is automatic, once percentile ranks are available. For example, the range of percentile ranks 40 through 59 becomes stanine 5. (See Figure 6.)

National norms, as described above, have meaning for all school systems, regardless of the ways in which a system's pupils may be different from the national population as a whole. This is true because national norms describe one reality, i.e., the typical performance of the nation's schoolchildren.

In this respect they are an important frame of reference. However, our test publisher may choose to provide other sets of norms for certain subgroups of the population. Thus, he may develop distributions of raw scores for pupils in large cities or for suburban areas, for geographical regions or for types of schools (Catholic school norms, for example). From these distributions, the publisher may develop percentile rank or stanine norms. It should be noted, however, that grade equivalent norms are never developed for any norm group other than the national school population.

The Manual: Information for the Test User

In selecting a standardized test series to fill his own testing needs, the prospective test user should look for assurances that the test is appropriate for his system and that, in fact, the norm sample population is "representative" of the total national school population. The test developer, in his manual for the test, will usually take care to describe the way in which the norm or standardization group was chosen and, as well, to describe that sample in terms of other criteria related to school performance. The test user should, moreover, look for evidence that the test itself is good for the purposes for which he intends to use it, since even "beautiful norms" are worthless if they are for a test with no value in his specific situation. The test publisher will probably collect the pertinent information and data into a manual for the test. Much of what the test user needs to evaluate the true worth of the test itself will typically be presented in sections labeled *validity* and *reliability*.

Validity may be described as the degree to which a test serves the purpose for which it is used; that is, how close it comes to doing what you want it to do. Since validity is specific to the purposes of the test, different kinds of validity evidence are appropriate to different kinds of tests. Most survey achievement tests are intended to assess current status in the subject areas being tested. They are in this sense, then, a sample set of items which are representative of a universe of information. The type of validity information most needed by the user is the blueprint or set of specifications on which the test was built and which has been described earlier. Test publishers will typically include some such content outline information in the manual. The user himself will then make his own determination of *content validity:* that is, the degree to which the test's specifications match his school's own objectives and curriculum. Included in the manual as well may be a discussion of the way in which the blueprint was developed, by whom, and other pertinent data.

The test developer cannot present statistical data to support a claim of content validity, since such validity must be specific to each school. Publishers may, however, present correlations of test scores with some other criterion (perhaps another well-known achievement test) as evidence of the validity or usefulness of the new test.

One of the most fundamental requirements of any measurement is that it be reliable; in other words, the obtained scores should be dependable. Test *reliability* may be described as the extent to which a test is consistent or stable in its measurement of whatever it is measuring. As with other measuring instruments, however, no test is perfectly accurate. The degree of accuracy possible depends in part on the fineness of the distinctions the test is expected to make. The greater the spread of scores the pupils have obtained, the more reliably the test will be able to distinguish among them. Since there is no one reliability figure for a test, the figures the test publisher presents are estimates of the test's reliability under typical circumstances. The publisher will also present, along with the reliability estimates, estimates of the "measurement error." In this way, the user may take into account the lack of perfect accuracy of the test when he interprets the scores obtained.

We noted, in the section describing the item analysis, the attention given to the degree to which questions distinguished between high- and low-scoring pupils. The total of the discrimination indices for questions chosen for final forms gives a crude indication of reliability and, thus, a preliminary assurance the standardized test will be fairly reliable.

At the time of standardization, several statistical procedures will be employed in analyzing the data and the results will be reported as evidence of the test's reliability. One statistic frequently reported is the "split-half" reliability coefficient which is obtained when the two halves of the test are correlated to provide an answer to the question, "How consistently is the test measuring the same thing throughout?" Another type of

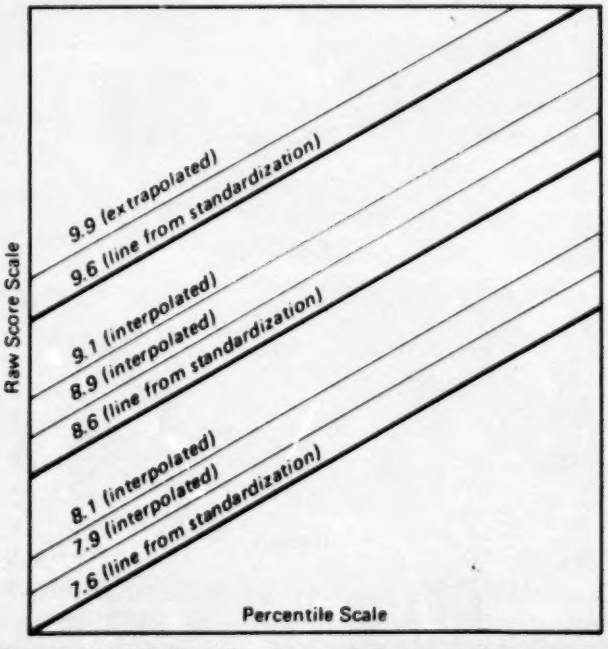


Figure 5. Schematic Illustration of Percentile Rank Norm Lines for Beginning, Middle, and End of Year

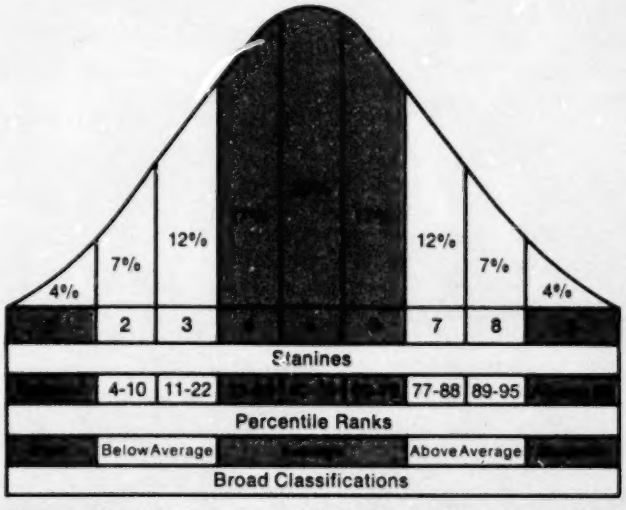


Figure 6. Schematic Illustration of a Normal Distribution of Scores, Showing the Relationship Between Percentile Ranks and Stanines

reliability statistic which may be presented is the "test-retest" or "alternate-form" reliability coefficient which is an indication of the correspondence between results on alternate forms and thus a measure of their reliability.

Since no test score is ever completely reliable, the test's developer may include data in his manual estimating the possible magnitude of error in the obtained scores of individuals. These data are called standard errors of measurement and are usually given for each test at each grade. He may also present data concerning the heterogeneity of his norm sample by giving the standard deviations of test scores of his norm groups, as well as their mean and median scores. When more than one form of a test is published, the test developer may provide evidence in the test manual that the alternate forms are comparable. As we have seen, the final forms are built to be as alike as possible in terms of item analysis data. However, the standardization program will probably either standardize each form directly or include a supplementary experimental program in which alternate forms will be administered to matched groups in order to relate scores on other forms to scores on the standardized form.

The test developer may include data concerning the intercorrelations among subtest scores in his achievement battery, the correlation of the achievement test scores with mental ability test scores, data concerning sex differences in test scores, and so on. Finally, the test publisher will usually give the test user suggestions for the interpretation and use of test scores by teachers, curriculum specialists, guidance counselors, and administrative personnel. These should be considered "must" reading for any test user who wishes to get the greatest possible benefit from his use of the test.

When all of these procedures have been completed and the results analyzed and interpreted and written up in a way the test user will find meaningful, the new standardized achievement test is ready for publication.

The paragraphs above have described only the building of a standardized achievement test. Of course, many other types of test instruments may be built and standardized by a test publisher, including tests of general mental ability or school learning potential, prognostic and diagnostic tests of various types, tests of specific aptitudes, interest and personality measures, etc. Many of the steps in developing these other types of tests are similar to those described above; in other ways, however, they may be quite different. In choosing the questions for a standardized scholastic aptitude test, for instance, emphasis would be placed, not on the content of the curriculum, but rather on skills or tasks which are found to be predictive of success in academic programs. In such cases, too, the type of validity sought is predictive, rather than content validity, and the estimates of reliability will be concerned more with stability over time than are those used in estimating the reliabilty of achievement measures.

Standardized tests are an important tool for evaluation of the educational process but they are not the only type of test which publishers build. The essence of a standardized test is in the use of a norm-reference: that is, that scores on the test are described in terms of a comparison with the scores others have made on the same test. The other major category of test instruments are those which use a criterion-reference rather than a norm-reference. Scores on criterion-referenced tests are described, not in terms of a comparison with what others have done, but in terms of the degree of attainment of some specific criterion or standard of achievement.

Because criterion-referenced and norm-referenced tests are built for different purposes and used in different ways, the techniques used to build them are also somewhat different. The present publication deals only with norm-referenced tests. A later publication will take up the question of how a criterion-referenced test is built.

The development of a good test instrument, of whatever type, is a time-consuming and demanding task. At Harcourt Brace Jovanovich, Inc., several years of research and development go into the publication of a major standardized test series. The painstaking care with which a published test has been developed, however, will help to assure that the resulting instrument will be of value to all those involved in the educational process.

TEST SERVICE NOTEBOOKS are issued from time to time as a professional service of The Psychological Corporation. Inquiries, comments, or requests for additional copies may be addressed to your Regional Office, or write Information and Advisory Service, The Psychological Corporation, New York, NY 10017.